



PROGRAMA INSTITUCIONAL DE BOLSAS DE INICIAÇÃO CIENTÍFICA PIBIC

RELATÓRIO FINAL

Predição de Séries Temporais Baseada em Redes Neurais Artificiais

Submetido à
Pró-Reitoria de Pesquisa da Universidade Estadual de Campinas

Departamento de Engenharia de Computação e Automação Industrial (DCA)
Faculdade de Engenharia Elétrica e de Computação (FEEC)
Universidade Estadual de Campinas (UNICAMP)
CEP 13083-852, Campinas - SP

Aluno: João Pedro de Oliveira Pagnan
Orientador: Prof. Levy Boccato

Campinas, 31 de agosto de 2021

1 Introdução

Depois de uma exposição sobre certos modelos preditores com redes neurais artificiais e sobre as características principais de sistemas com dinâmica caótica, apresentados no relatório parcial (seções 2.1 e 2.2, respectivamente), este relatório final irá expor os cenários utilizados para a análise, a metodologia utilizada, os resultados obtidos e, por fim, as conclusões desta pesquisa sobre o desempenho de redes neurais artificiais na predição de séries temporais originadas por sistemas com dinâmica caótica, além de alguns detalhes sobre as arquiteturas estudadas nesta segunda parte da pesquisa, que foram a *Gated Recurrent Unit* (GRU) [1] e a *Echo State Network* (ESN) [2].

A seção 2 apresenta os quatro cenários escolhidos para a análise do desempenho das redes neurais, sendo dois destes a tempo discreto e dois a tempo contínuo. No caso, os sistemas a tempo discreto foram o mapa de Hénon [3] e o mapa logístico [4]. Já os sistemas a tempo contínuo foram o sistema de Lorenz [5] e as equações de Mackey-Glass [6].

Já na seção 3, discutiremos os dois modelos previamente citados que foram estudados e implementados, juntamente com as redes neurais apresentadas no relatório parcial, nesta segunda parte da pesquisa.

A seção 4 realiza uma exposição da busca em grade [7] realizada em cada modelo, expondo quais foram os parâmetros testados e os critérios definidos para o processo de busca. Além disso, a seção 4 também apresentará a metodologia utilizada para definir o número de amostras de entrada de cada modelo preditor (nesse caso, chamado de K), além de indicar qual foi a progressão do erro quadrático médio (EQM) em função do valor de K para cada modelo nos quatro cenários.

Por fim, as seções 5 e 6 apresentam os resultados e as conclusões obtidas, respectivamente, encerrando, assim, esta pesquisa de iniciação científica.

2 Cenários escolhidos

Antes de falarmos sobre os cenários utilizados na análise, vale a pena reforçarmos as características principais de sistemas com dinâmica caótica.

Sistemas caóticos se destacam pois, apesar de serem determinísticos, apresentam dependência sensível em relação às condições iniciais (DSCI). Dessa forma, duas trajetórias que partem de posições relativamente próximas no espaço de estados podem evoluir de uma forma totalmente distinta devido às não-linearidades presentes que amplificam as diferenças entre essas condições iniciais [8].

De forma resumida, a dinâmica caótica é marcada pela presença dos seguintes aspectos [9]:

1. Forte sensibilidade com respeito às condições iniciais;

2. A evolução temporal das variáveis de estado (parâmetros de ordem do sistema) é rápida e tem uma aparência errática;
3. Um sinal originado por um sistema caótico tem espectro de potências contínuo e de faixa larga;
4. Há uma produção de informação por parte do sistema;
5. Dão origem a atratores estranhos (estruturas topológicas que ditam a evolução temporal do fluxo de um sistema caótico) [10].

Retomados os pontos principais da dinâmica caótica, daremos continuidade à discussão apresentando os cenários escolhidos para a análise. Vale mencionar que, na simulação numérica dos quatro sistemas, foram geradas 5000 amostras para cada série temporal. Além disso, nos sistemas multidimensionais, como o mapa de Hénon e o sistema de Lorenz, consideramos apenas a variável de estado x na previsão, em ambos os casos.

2.1 Sistema de Lorenz

O sistema de Lorenz foi um dos sistemas dinâmicos caóticos a tempo contínuo abordados nessa pesquisa. Este sistema foi um dos primeiros grandes trabalhos envolvendo a noção de regime caótico, sendo considerado por muitos a pesquisa que inaugurou a área [11].

Através de simulações numéricas de uma dinâmica atmosférica, o matemático e meteorologista Edward Norton Lorenz observou uma dependência sensível às condições iniciais em certos sistemas dinâmicos [8]. Dando continuidade à seus experimentos, Lorenz modelou, através de três equações diferenciais, o fluxo de um fluido em um volume uniformemente aquecido na camada inferior e uniformemente resfriado na camada superior [5]:

$$\frac{dx}{dt} = -\sigma \cdot (x - y) \quad (1a)$$

$$\frac{dy}{dt} = x \cdot (\rho - z) - y \quad (1b)$$

$$\frac{dz}{dt} = x \cdot y - \beta \cdot z \quad (1c)$$

sendo σ , ρ e β constantes reais, estando relacionadas a certas características físicas do sistema, como o número de Prandtl, o número de Rayleigh e as dimensões do volume que o fluido ocupa [8].

Utilizando $\sigma = 10$, $\rho = 28$ e $\beta = 8/3$, Lorenz demonstrou que esse sistema de equações diferenciais exibe comportamento caótico, sendo que a maioria das condições iniciais $[x(0) \ y(0) \ z(0)] = [0 \ 0 \ 0]^T$ convergem para um atrator estranho (nesse caso, atrator de Lorenz).

A figura 1 indica a série temporal em x , que foi utilizada em nossa análise, para $[x(0) \ y(0) \ z(0)]^T = [0.1 \ 0 \ 0]^T$, e o atrator de Lorenz para a trajetória. Para a simulação,

os parâmetros do sistema foram configurados para os mesmos valores utilizados por Lorenz (exibidos no parágrafo anterior), e foi utilizado $dt = 0.01$ para resolver as equações diferenciais numericamente.

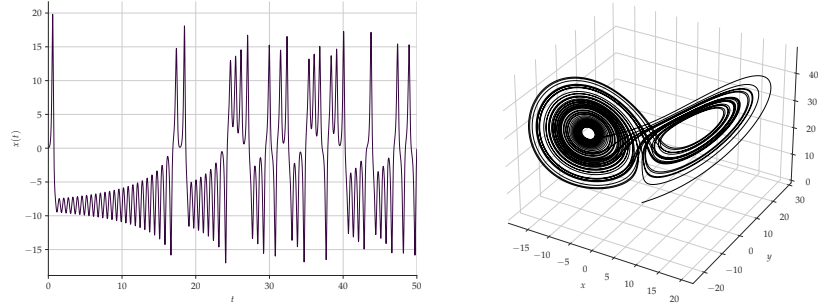


Figura 1: À esquerda, a série temporal em x do sistema de Lorenz simulado e, à direita, o diagrama de fases correspondente à simulação

2.2 Mapa de Hénon

O mapa de Hénon foi um dos sistemas a tempo discreto escolhidos para esta pesquisa. Esse sistema foi proposto pelo astrônomo e matemático francês Michel Hénon em 1976 como um modelo simplificado de uma seção de Poincaré do atrator de Lorenz, sendo descrito pelas equações abaixo [3]:

$$x[n + 1] = y[n] + 1 - a \cdot (x[n])^2 \quad (2a)$$

$$y[n + 1] = b \cdot x[n] \quad (2b)$$

A presença de dinâmica caótica neste sistema discreto irá depender dos valores dos parâmetros a e b . Hénon mostrou, em sua pesquisa, que, para $a = 1.4$ e $b = 0.3$, há a presença de um atrator estranho do diagrama de fases desse sistema dinâmico discreto.

O mapa de Hénon realiza um mapeamento de dois pontos, chamados de pontos fixos. Para os valores dos parâmetros a e b mencionados anteriormente, tais pontos são dados por:

$$x = \frac{\sqrt{609} - 7}{28} \approx 0.631354477$$

$$y = \frac{3(\sqrt{609} - 7)}{280} \approx 0.189406343$$

Um desses pontos está sobre o atrator e é instável. Tal instabilidade é confirmada quando é realizada uma análise da trajetória com outros pontos próximos a este, onde percebe-se que, dependendo da região do atrator pela qual o ponto em análise se aproxima ao ponto fixo, a trajetória irá divergir ou convergir para o ponto no atrator.

A figura 2 mostra a série temporal referente à variável x e o atrator obtido com a simulação para $[x[0] \ y[0]]^T = [1 \ 0]^T$.

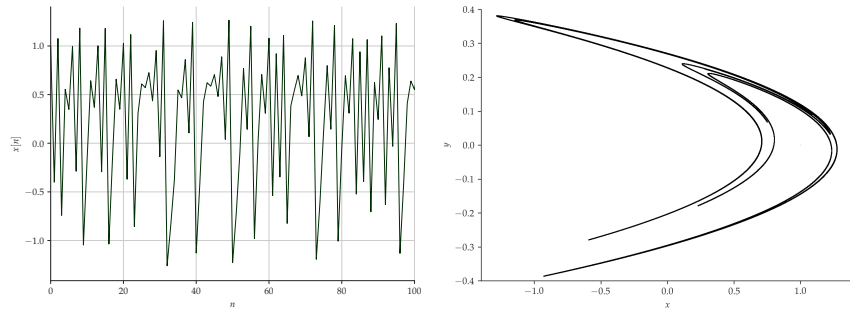


Figura 2: À esquerda, as cem primeiras iterações da série temporal em x do mapa de Hénon e, à direita, o atrator correspondente à simulação

2.3 Mapa logístico

Descrito em 1976 pelo biólogo e matemático Robert May, o mapa logístico representa uma das formas de modelar a população de uma determinada espécie em certos instantes de tempo [4]. A equação a diferenças que descreve esse sistema pode ser vista abaixo:

$$x[n + 1] = r \cdot x[n] \cdot (1 - x[n]) \quad (3)$$

sendo x_n um número real entre 0 e 1 que representa a razão entre o tamanho atual da população e o tamanho máximo desta, e tendo r um valor real entre 0 e 4, representando a taxa de crescimento desta população.

Nesse caso, o sistema não chega a operar em caos para qualquer valor de r . Robert May, através de análises numéricas [8], viu que, para valores maiores que 3.44949, o sistema entra em regime caótico. Interessantemente, ao aumentar mais o valor do parâmetro r , o cientista percebeu que há certas zonas de estabilidade, onde o sistema converge para um ciclo-limite.

Inicialmente, após sair do regime caótico, a dinâmica terá período 2, e, com o aumento de r , o sistema começa a ter período 4, 8, 16, 32, e assim por diante, até voltar, novamente, ao movimento caótico. Esse padrão de ir do caos a estabilidade para, então, voltar ao regime caótico, se repete até que r atinja seu valor máximo.

Quando o físico e matemático Mitchell Feigenbaum estudou esse fenômeno, também presente em outros sistemas caóticos e em certas estruturas fractais, percebeu que a razão entre o comprimento de dois intervalos sucessivos de bifurcação tende a $\delta \approx 4.66920$ [12]. Tal valor tem a alcunha de constante de Feigenbaum, não estando relacionado a nenhuma outra propriedade matemática ou constante da natureza conhecida até então.

Como o estudo visa analisar o desempenho para sistemas caóticos, foi utilizado $r = 3.86$, que, conforme será visto no diagrama de bifurcação abaixo, faz com que a série temporal dada pela equação (3) opere em caos.

A figura 3 indica a série temporal obtida partindo de $x[0] = 0.5$ e o diagrama de bifurcação para este sistema de equações.

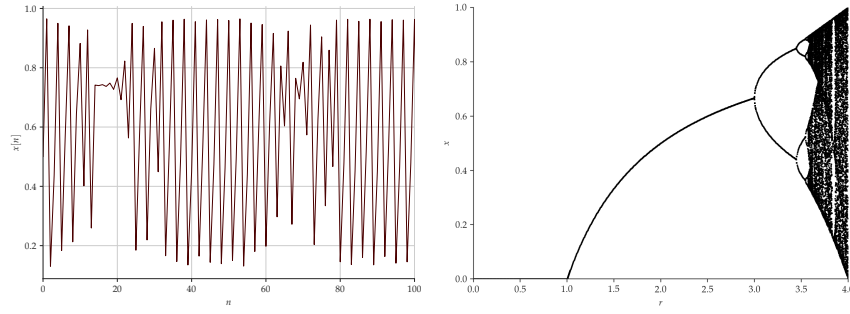


Figura 3: À esquerda, as cem primeiras iterações da série temporal do mapa logístico e, à direita, o diagrama de bifurcação deste sistema

2.4 Equações de Mackey-Glass

Por fim, o último sistema caótico simulado, também contínuo, está associado às equações de Mackey-Glass, as quais modelam o controle hormonal da produção de células brancas do sangue e podem ser vistas abaixo [6]:

$$\frac{dP(t)}{dt} = \frac{\beta_0 \cdot \theta^n}{\theta^n + P(t - \tau)^n} - \gamma \cdot P(t) \quad (4a)$$

$$\frac{dP(t)}{dt} = \frac{\beta_0 \cdot \theta^n \cdot P(t - \tau)}{\theta^n + P(t - \tau)^n} - \gamma \cdot P(t) \quad (4b)$$

sendo $P(t)$ a densidade de tais células em um instante de tempo e β_0, θ, n, τ e γ valores reais relacionados a certos parâmetros hormonais de um organismo, geralmente sendo determinados experimentalmente.

Neste caso, conforme demonstrado em [6], a equação (4b) exhibe comportamento caótico para valores mais altos de τ . Para a simulação numérica, foi utilizado $n = 10$, $\gamma = 0.1$, $\beta = 0.2$, $\theta = 1$, $\tau = 22$, $dt = 1.0$ e $P(0^-) = 0.1$, gerando novamente 5000 amostras. A série e o atrator obtidos podem ser vistos na figura 4.

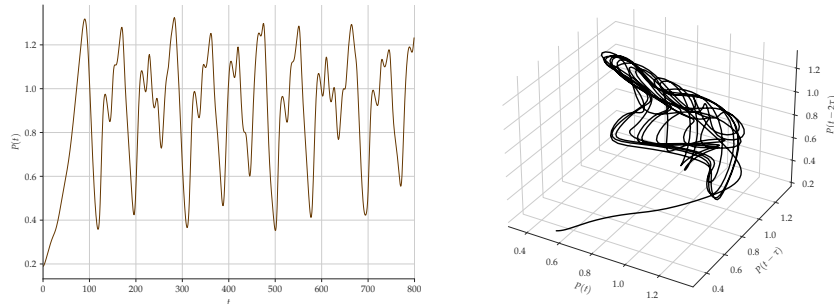


Figura 4: À esquerda, a série temporal da equação (4b) exibida de $t = 0$ a $t = 800$ e, à direita, o atrator correspondente à simulação

Curiosamente, os pesquisadores perceberam que regimes caóticos dessa equação estão correlacionados com certos problemas fisiológicos nos nossos organismos [6].

Com isso, finalizamos a exposição sobre os cenários de sistemas com dinâmica caótica escolhidos para a análise comparativa. Na próxima seção, falaremos a respeito dos dois modelos estudados e implementados nesta segunda parte da pesquisa. Em seguida, apresentaremos a metodologia aplicada nesses e nos outros modelos mostrados no relatório parcial, assim como as particularidades para cada caso.

3 Modelos estudados

Como o estudo de redes neurais artificiais foi iniciado na primeira metade da pesquisa, uma exposição mais detalhada dos fundamentos destes modelos (no caso, as redes MLP e as rede recorrentes LSTM) pode ser vista no relatório parcial.

Assim, as próximas duas subseções abordarão especificamente as redes recorrentes GRU e ESN.

3.1 Gated Recurrent Unit (GRU)

Como já mencionamos anteriormente, as redes recorrentes *Gated Recurrent Unit* possuem células computacionais semelhantes às células das redes LSTM, apresentadas no relatório parcial. Sua estrutura interna pode ser vista na figura 5, juntamente com as equações que a descrevem.

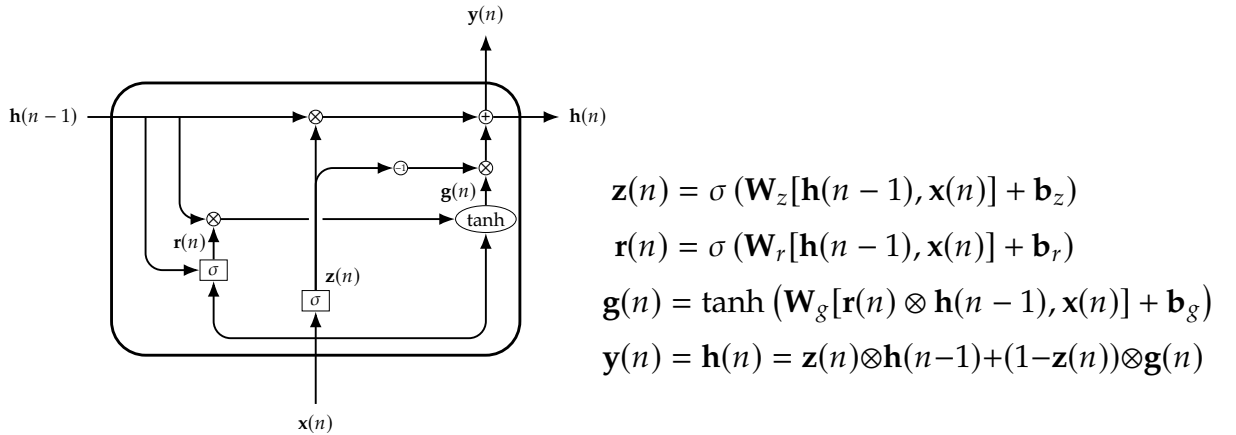


Figura 5: Estrutura e equações de uma célula GRU

Primeiramente, por ser uma rede recorrente, há uma realimentação nestas estruturas. Ou seja, a saída atual é determinada levando em conta estados passados da rede.

Mas, diferentemente da LSTM, na célula GRU há somente um vetor de estados passados $\mathbf{h}(n)$, e o vetor $\mathbf{z}(n)$ determinará se a memória $\mathbf{h}(n-1)$ será mantida ou apagada. Além disso, também é utilizado o vetor $\mathbf{r}(n)$, que controla qual parte de $\mathbf{h}(n-1)$ irá compor $\mathbf{g}(n)$.

Intuitivamente, pelo fato das GRU serem, basicamente, uma versão mais simples das redes LSTM, espera-se que a última seja melhor que a primeira. Porém, conforme mostrado em [1], em determinadas aplicações a GRU é superior à LSTM. Além disso, a implementação de um modelo preditor com a rede GRU é bem similar a implementação de um preditor baseado na LSTM. Por causa disso, incluímos a *Gated Recurrent Unit* na análise comparativa.

O outro modelo recorrente estudado nesta segunda metade da pesquisa foi as redes com estados de eco. Essas arquiteturas são relativamente distintas das outras redes neurais estudadas. Veremos então o que diferencia esse modelo dos outros na próxima subseção.

3.2 Echo State Network (ESN)

Diferentemente das redes neurais mostradas até aqui, incluindo as citadas no relatório parcial, as redes com estados de eco baseiam-se na teoria de computação por reservatório, chamada assim pois é utilizado um reservatório de comportamentos dinâmicos que, ao serem combinados linearmente nos neurônios da camada de saída, produzem as respostas da rede aos estímulos de entrada [13].

A figura 6 apresenta a estrutura interna de uma rede ESN.

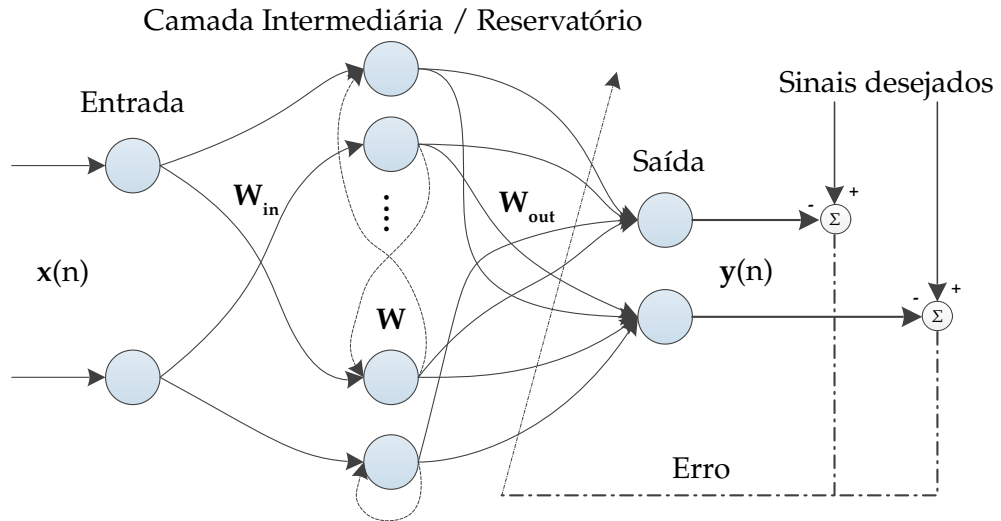


Figura 6: Estrutura típica de uma rede ESN (figura adaptada de [13])

As equações em (5) descrevem a ESN, sendo $\mathbf{x}(n)$ o vetor de entrada, $\mathbf{W}_{in} \in \mathbb{R}^{N \times K}$ os pesos da camada de entrada, $\mathbf{u}(n)$ o vetor que representa as ativações dos N neurônios não-lineares totalmente conectados do reservatório, representando os estados da rede, $\mathbf{W} \in \mathbb{R}^{N \times N}$ os pesos das conexões recorrentes do reservatório, $\mathbf{f}(\cdot)$ denota as funções de ativação das unidades internas da rede, $\mathbf{W}_{out} \in \mathbb{R}^{L \times N}$ é a matriz dos pesos da camada de saída e, por fim, $\mathbf{y}(n)$ denota as saídas da rede.

$$\mathbf{u}(n + 1) = \mathbf{f}(\mathbf{W}_{in} \cdot \mathbf{x}(n + 1) + \mathbf{W} \cdot \mathbf{u}(n)) \quad (5a)$$

$$\mathbf{y}(n + 1) = \mathbf{W}_{out} \cdot \mathbf{u}(n + 1) \quad (5b)$$

O grande diferencial prático desta rede é que os pesos das conexões recorrentes dentro do reservatório \mathbf{W} são ajustados com valores fixos antes do treinamento da camada de saída. Esses parâmetros são determinados tendo em vista a propriedade de estados de eco, existindo alguns métodos simples para a criação aleatória dos pesos e que asseguram a propriedade, conforme demonstrado em [2].

Devido a isso, o ajuste dos neurônios na camada de saída pode ser realizado com um procedimento de regressão linear, sem necessitar de um treinamento iterativo, como nas outras redes recorrentes. Esse fato economiza tempo de processamento e garante que será atingido o mínimo global da função custo utilizada.

Com isso, finalizamos o restante da exposição teórica desta pesquisa. A próxima seção descreverá a metodologia utilizada nesta análise.

4 Metodologia

Nesta segunda parte da pesquisa de iniciação científica foram implementados modelos preditores com as quatro redes neurais artificiais estudadas (MLP, LSTM, GRU e ESN) para os quatro cenários exibidos na seção 2.

Iniciaremos a discussão com a apresentação da busca em grade realizada em todos os modelos, de forma a determinar os parâmetros ótimos das redes neurais em cada cenário. Em seguida, utilizando os melhores parâmetros, avaliamos a progressão do erro quadrático médio em função do número de amostras de entrada K do modelo preditor. Por fim, comparamos qual foi a média e o desvio padrão do EQM com o melhor valor de K de cada modelo nos quatro cenários. Em todos os experimentos, o horizonte de predição utilizado foi $L = 3$. Assim, iremos prever o valor da terceira iteração à frente do valor atual da série temporal.

Vale reforçar que, para todos os modelos e cenários, os dados de treinamento correspondiam a cerca de 85% das amostras das séries temporais geradas, que correspondiam aos primeiros 4250 valores. Além disso, o conjunto de validação, utilizado na busca em grade para avaliar a predição de uma configuração testada, corresponde a 10% dos dados de treinamento (as últimas 425 amostras). Assim, os dados de treinamento efetivamente correspondem às primeiras 3825 amostras, os dados de validação correspondem às 425 amostras depois destas e, por fim, o conjunto de teste corresponde aos últimos 750 dados. Essa proporção entre os dados de treinamento, de teste e de validação foi utilizada para todas as redes analisadas.

Em seguida, na seção 4.2, será apresentado a análise feita

4.1 Configurações utilizadas

Conforme dito anteriormente, foi realizado uma busca em grade com validação cruzada para determinar as melhores configurações de cada rede neural para a predição das séries temporais em cada cenário. Para cada arquitetura, um conjunto de valores candidatos foi gerado para cada hiperparâmetro e todas as combinações possíveis foram avaliadas tendo em vista o desempenho do modelo em dados de validação.

Além da diferença no que será testado, houveram algumas distinções entre o processo realizado na MLP e nas redes recorrentes. Como a relação temporal entre as amostras não é importante para a MLP, a busca em grade considerou uma validação cruzada com 4 *fold*s nos dados de treinamento. Ou seja, para cada configuração, foi avaliado o desempenho desta rede na predição de quatro subconjuntos das quatro séries temporais escolhidas.

Já nas redes recorrentes, foi considerado um esquema de validação do tipo *holdout*. Esse processo dividiu o conjunto de treinamento (novamente composto por 85% dos dados gerados) em 4 seções. Cada seção era composta por uma fração dos dados de treinamento, sendo que cada seção incluía a seção anterior no seu conjunto de dados. Por exemplo, a segunda seção obtida pelo *holdout* inclui a primeira seção e mais algumas amostras, além de uma subdivisão de validação que será utilizada para avaliar o resultado. Com isso, obtém-se conjuntos sequenciais de dados para avaliação do desempenho. Esse procedimento é necessário para as redes recorrentes, pois a relação temporal entre os dados de entrada deve ser levada em conta.

Por fim, para a obtenção dos melhores hiperparâmetros, foi considerado que $K = 4$.

Com esses detalhes expostos, iniciaremos a exposição com a busca em grade realizada na rede MLP para, em seguida, falarmos sobre as configurações obtidas para as redes recorrentes.

4.1.1 MLP

Na busca em grade para a MLP, avaliamos diferentes tamanhos de *batch* (2, 4, 8, 16, 32), se haveria ou não uma camada de *batch normalization* na entrada da rede MLP, diferentes opções de função de ativação dos neurônios na camada intermediária (SELU, ReLU, ELU, tanh, Sigmoid), esquemas alternativos para a inicialização dos pesos (Glorot, He, Lecun, tanto com distribuição normal, quanto uniforme), diferentes quantidades de neurônios na camada intermediária (5, 10, 15, 20, 30, 50, 75, 100) e, também, diferentes valores para a taxa de aprendizagem (0.001, 0.003, 0.005, 0.008, 0.01).

Vale reforçar que foi utilizada apenas uma camada intermediária em cada cenário. As melhores configurações obtidas podem ser vistas na tabela 1.

Cenário	<i>Batch normalization</i>	Tamanho do <i>batch</i>	Ativação	Inicialização	Nº de neurônios	Taxa de aprendizagem
Mapa de Hénon	Não	8	Sigmoid	Glorot Normal	50	0.003
Mapa logístico	Não	2	tanh	Glorot Uniforme	10	0.003
Sistema de Lorenz	Não	2	SELU	Lecun Normal	50	0.001
Equações de Mackey-Glass	Não	4	tanh	Glorot Normal	5	0.001

Tabela 1: Melhores parâmetros para a rede MLP nos cenários em análise

4.1.2 LSTM e GRU

Para as redes recorrentes LSTM e GRU, foi avaliado o tamanho do *batch* (2, 4, 8, 16, 32), a inicialização dos pesos (Glorot Uniforme, Glorot Normal), o número de neurônios recorrentes na camada intermediária (5, 10, 15, 20, 30, 50, 75, 100), e a taxa de aprendizagem (0.001, 0.003, 0.005, 0.008, 0.01), novamente utilizando apenas uma camada intermediária e mantendo a função de ativação usual da célula recorrente (tanh). Por conta disso, para o sistema de Lorenz, foi feito um ajuste de escala para evitar a saturação dos neurônios.

As tabelas 2 e 3 exibem as melhores configurações obtidas em cada cenário para as redes LSTM e GRU, respectivamente.

Cenário	Tamanho do <i>batch</i>	Inicialização	Nº de neurônios	Taxa de aprendizagem
Mapa de Hénon	4	Glorot Normal	15	0.005
Mapa logístico	2	Glorot Uniforme	100	0.008
Sistema de Lorenz	4	Glorot Uniforme	15	0.003
Equações de Mackey-Glass	2	Glorot Uniforme	50	0.003

Tabela 2: Melhores parâmetros para a rede LSTM nos cenários em análise

Cenário	Tamanho do <i>batch</i>	Inicialização	Nº de neurônios	Taxa de aprendizagem
Mapa de Hénon	4	Glorot Normal	30	0.003
Mapa logístico	2	Glorot Normal	100	0.003
Sistema de Lorenz	8	Glorot Uniforme	30	0.001
Equações de Mackey-Glass	2	Glorot Uniforme	10	0.005

Tabela 3: Melhores parâmetros para a rede GRU nos cenários em análise

4.1.3 ESN

Para esta rede foi realizada uma busca em grade para determinar o número de neurônios no reservatório (30, 50, 70, 90, 100, 120, 140, 160, 180, 200, 240, 280, 320, 360, 400, 440, 480, 500) e o valor para o raio espectral (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.96, 0.97, 0.98, 0.99), que determina o maior valor singular da matriz de pesos do reservatório. Por tratar-se de uma rede recorrente, também foi utilizado o processo de *holdout* descrito na seção 4.1. Vale também mencionar que a taxa de vazamento, que representa a velocidade com a qual o reservatório atualiza suas dinâmicas, foi fixada em 0.9 para todos os cenários.

A tabela 4 exibe as melhores configurações para a ESN em cada um dos cenários.

Parâmetro	Mapa de Hénon	Mapa logístico	Sistema de Lorenz	Equações de Mackey-Glass
Nº de neurônios	500	500	120	500
Raio espectral	0.1	0.1	0.2	0.4

Tabela 4: Melhores parâmetros para a rede ESN nos cenários em análise

Com isso, concluímos a apresentação das melhores configurações para as redes neurais avaliadas em todos os cenários escolhidos. Na próxima seção, falaremos sobre outra análise realizada, dessa vez não sobre os modelos em teste, mas sim sobre certas características das séries temporais utilizadas.

4.2 Estudo da autocorrelação parcial das séries temporais

Com as melhores configurações para cada rede e cenário obtidas, foi analisada a progressão do erro quadrático médio para cada valor de K . A faixa de valores para K a ser testada foi determinada utilizando a autocorrelação parcial das séries temporais de cada

sistema analisado, levando em consideração os valores de K que têm as autocorrelações parciais mais relevantes.

A figura 7 mostra a autocorrelação parcial para as séries temporais utilizadas.

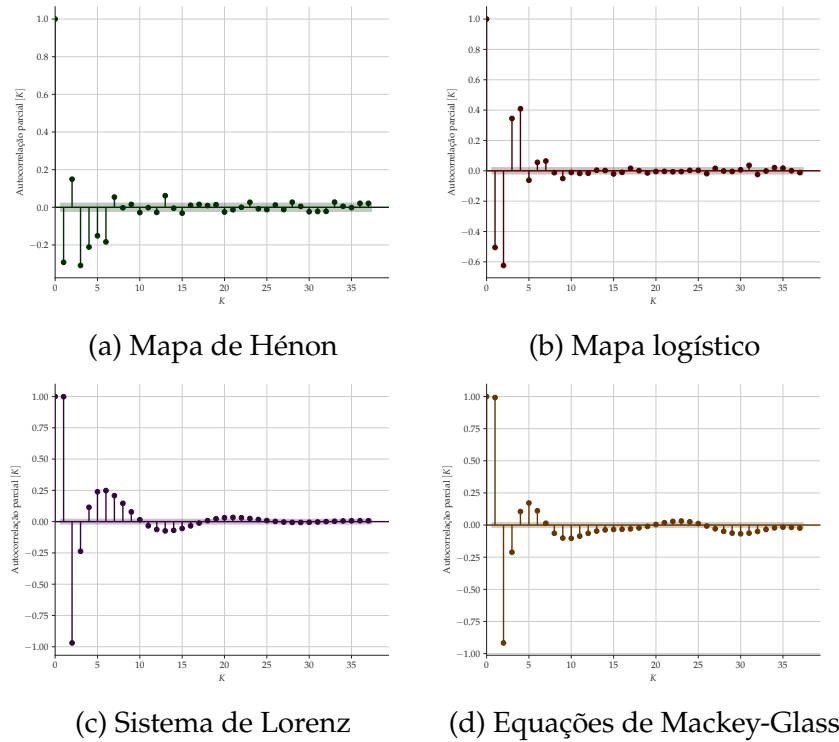


Figura 7: Autocorrelação parcial para as séries temporais utilizadas

Vale mencionar que, a autocorrelação parcial encontra a correlação dos resíduos (chamados assim pois são mantidos após a remoção dos efeitos já explicados pelos primeiros atrasos) com o próximo valor de atraso, removendo correlações entre este e atrasos passados. Logo, essa medida descreve somente a relação direta entre duas observações da série afastadas por K instantes de tempo.

A tabela 5 mostra a faixa de valores para K a ser testada obtida com a análise descrita nos parágrafos anteriores.

	Mapa de Hénon	Mapa logístico	Sistema de Lorenz	Equações de Mackey-Glass
Faixa de valores	1 ~ 9	1 ~ 8	1 ~ 10	1 ~ 7

Tabela 5: Faixa de valores inteiros para K a ser testada em cada cenário para todas as redes

4.3 Análise do melhor valor para K

Com as melhores configurações para cada rede e cenário obtidas, foi analisada a progressão do erro quadrático médio para cada valor de K . A faixa de valores para K a ser testada foi determinada utilizando a autocorrelação parcial das séries temporais de cada sistema analisado, levando em consideração os valores de K que têm as autocorrelações parciais mais relevantes.

Para realizar esse procedimento, cada rede (com as configurações ótimas) foi treinada utilizando 85% dos dados gerados, sendo que 10% dos dados de treinamento foi utilizado como o conjunto de validação (nas redes MLP, LSTM e GRU). Em seguida, com o modelo treinado, foi avaliado o EQM no conjunto de teste (que corresponde às últimas 750 amostras). Esse processo foi realizado 5 vezes para cada K , obtendo assim um valor médio e desvio-padrão para cada modelo e em cada cenário.

Vale reforçar que, como o ajuste de parâmetros da ESN possui solução em forma fechada, não foi necessário utilizar um conjunto de validação no treinamento. No caso da MLP, LSTM e GRU, o processo iterativo de ajuste dos pesos utilizou um conjunto de validação de forma a seguir o procedimento de *early stopping* para evitar o sobreajuste da rede.

A figura 8 mostra a progressão do EQM observada em cada cenário para cada uma das redes estudadas.

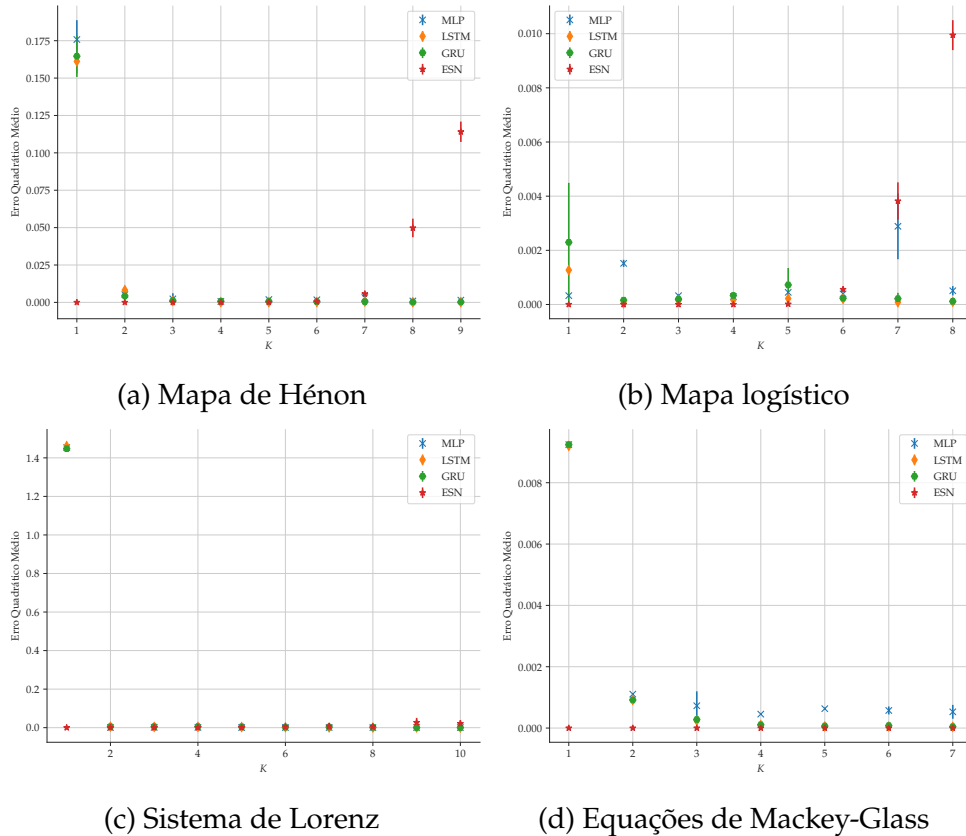


Figura 8: Progressão do erro quadrático médio para cada valor de K nos cenários avaliados para todas as redes neurais testadas

Nas tabelas 6a, ??, ?? e ??, constam os valores médios e desvios-padrões para cada valor K dos modelos avaliados, no cenário do mapa de Hénon, do mapa logístico, do sistema de Lorenz, e das equações de Mackey-Glass, respectivamente.

K	MLP	LSTM	GRU	ESN
1	$(1.758 \pm 0.130) \times 10^{-1}$	$(1.612 \pm 0.081) \times 10^{-1}$	$(1.646 \pm 0.139) \times 10^{-1}$	$(4.657 \pm 3.054) \times 10^{-9}$
2	$(5.190 \pm 1.620) \times 10^{-3}$	$(8.223 \pm 1.426) \times 10^{-3}$	$(4.315 \pm 1.958) \times 10^{-3}$	$(9.764 \pm 3.727) \times 10^{-9}$
3	$(2.346 \pm 3.765) \times 10^{-3}$	$(9.028 \pm 7.506) \times 10^{-4}$	$(1.252 \pm 1.082) \times 10^{-3}$	$(8.699 \pm 2.645) \times 10^{-8}$
4	$(6.415 \pm 4.726) \times 10^{-4}$	$(2.498 \pm 0.854) \times 10^{-4}$	$(7.557 \pm 2.607) \times 10^{-4}$	$(1.595 \pm 0.425) \times 10^{-6}$
5	$(1.892 \pm 0.214) \times 10^{-3}$	$(2.256 \pm 1.282) \times 10^{-4}$	$(6.827 \pm 3.646) \times 10^{-4}$	$(3.355 \pm 0.519) \times 10^{-5}$
6	$(1.739 \pm 1.251) \times 10^{-3}$	$(2.341 \pm 0.471) \times 10^{-4}$	$(5.686 \pm 2.254) \times 10^{-4}$	$(5.113 \pm 1.617) \times 10^{-4}$
7	$(3.239 \pm 1.034) \times 10^{-3}$	$(4.771 \pm 2.439) \times 10^{-4}$	$(4.486 \pm 0.877) \times 10^{-4}$	$(5.731 \pm 0.882) \times 10^{-3}$
8	$(9.436 \pm 2.631) \times 10^{-4}$	$(1.578 \pm 0.689) \times 10^{-4}$	$(2.411 \pm 1.391) \times 10^{-4}$	$(4.978 \pm 0.620) \times 10^{-2}$
9	$(1.514 \pm 0.805) \times 10^{-3}$	$(2.151 \pm 0.605) \times 10^{-4}$	$(3.007 \pm 2.156) \times 10^{-4}$	$(1.141 \pm 0.068) \times 10^{-2}$

(a) Mapa de Hénon

Tabela 6: Valores médios e desvios-padrões do erro quadrático médio obtidos para cada K nos cenários utilizadas

5 Resultados

6 Análise e Conclusão

Analisando os resultados obtidos, percebe-se que, com exceção do cenário do sistema de Lorenz, a rede MLP foi consideravelmente pior do que as redes recorrentes. Também percebe-se que, dentre as redes recorrentes, a ESN obteve um EQM bem inferior ao obtido pela LSTM e pela GRU.

O desempenho inferior da rede MLP com relação às redes recorrentes provavelmente decorre do fato de que a relação temporal presente na LSTM, GRU e ESN auxilia na modelagem da dinâmica da série temporal. Já o pior desempenho da LSTM na série temporal do sistema de Lorenz provavelmente está relacionado aos efeitos estocásticos presentes no ajuste dos pesos sinápticos dessa rede neural que, conforme indicado em [14], é uma dificuldade em seu treinamento.

Algo interessante de mencionar é que, no geral, os desempenhos de todos os modelos estudados foram consideravelmente bons nas séries temporais dos sistemas a tempo contínuo (sistema de Lorenz e equações de Mackey-Glass), sendo que as diferenças foram mais pronunciadas na série temporal do mapa de Hénon e do mapa logístico. Provavelmente, a suavidade presente nas séries de Lorenz e de Mackey-Glass facilita a

modelagem do preditor, além do fato de que as séries temporais mencionadas são bem menos erráticas se comparadas às séries de Hénon e do mapa logístico.

O principal resultado observado foi que o modelo preditor utilizando a rede com estados de eco obteve um desempenho bem superior aos outros modelos, em todos os cenários. O erro quadrático médio foi tão baixo que, observando a predição nos dados de teste, praticamente não há diferença entre os valores reais e os valores previstos. Essa superioridade da ESN também é realçada na imagem comparativa do EQM, que indica que a ESN atingiu patamares de erro cerca de 100 ou até 10000 vezes menores que os outros modelos.

Esse resultado, aliado ao fato de que o treinamento da ESN é bem menos custoso computacionalmente se comparado ao das outras redes, mostra que esta rede é uma boa alternativa para futuros estudos de modelos preditores de séries temporais. Além disso, conforme mostrado em outros trabalhos como [15, 2, 13], a ESN também é uma boa ferramenta para outras tarefas de extração de informação, como em equalização de canais e separação de fontes.

Como sugestão de trabalhos futuros, pode ser avaliada a eficácia da ESN em reconstruir atratores através de séries temporais experimentais de sistemas caóticos. Se o desempenho para essa tarefa for tão bom quanto o obtido na predição das séries estudadas, a ESN pode tornar-se uma ferramenta ainda mais poderosa para a modelagem de sistemas com dinâmica caótica.

Referências

- [1] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.
- [2] H. Jaeger, "Echo state network," *scholarpedia*, vol. 2, no. 9, p. 2330, 2007.
- [3] M. Hénon, "A two-dimensional mapping with a strange attractor," *Communications in Mathematical Physics*, vol. 50, pp. 69–77, feb 1976.
- [4] R. M. May, "Simple mathematical models with very complicated dynamics," *Nature*, vol. 261, pp. 459–467, jun 1976.
- [5] E. N. Lorenz, "Deterministic nonperiodic flow," *Journal of atmospheric sciences*, vol. 20, no. 2, pp. 130–141, 1963.
- [6] M. C. Mackey and L. Glass, "Oscillation and chaos in physiological control systems," *Science*, vol. 197, no. 4300, pp. 287–289, 1977.
- [7] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, 2019.
- [8] N. Fiedler-Ferrara and C. P. C. do Prado, *Caos: uma introdução*. Editora Blucher, 1994.
- [9] R. R. de Faissol Attux, "Sobre dinâmica caótica e convergência em algoritmos de equalização autodidata," dissertação (mestrado), Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação, Campinas, SP, 2001.
- [10] D. Ruelle and F. Takens, "On the nature of turbulence," *Les rencontres physiciens-mathématiciens de Strasbourg-RCP25*, vol. 12, pp. 1–44, 1971.
- [11] J. Gleick, *Chaos: The amazing science of the unpredictable*. Vintage Publishing, 1998.

- [12] M. J. Feigenbaum, "Quantitative universality for a class of nonlinear transformations," *Journal of statistical physics*, vol. 19, no. 1, pp. 25–52, 1978.
- [13] L. Boccato, *Novas propostas e aplicações de redes neurais com estados de eco*. Tese (doutorado), Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação, Campinas, SP, 2013.
- [14] K. Doya *et al.*, "Bifurcations in the learning of recurrent neural networks 3," *learning (RTRL)*, vol. 3, p. 17, 1992.
- [15] H. Jaeger and H. Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication," *science*, vol. 304, no. 5667, pp. 78–80, 2004.